

**IN THE SPECIFICATION:**

(1) Please replace paragraph 130 on pages 55-56 with the following paragraph showing changes.

A critical input parameter to the compression algorithms is the error tolerance for numeric attribute  $X_i$  is specified as a percentage of the width of the range of  $X_i$ -values in the table. Another important parameter to the table compressor 100 is the size of the sample that is used to select the CaRT models in the final compressed table. For these two parameters, the default values of 1% (for error tolerance) and 50KB (for sample size), respectively, are used in all of the referenced experiments. Note that 50KB corresponds to 0.065% to 0.475% and .174% of the total size of the Forest-cover, Corel and Census data sets, respectively. Finally, unless stated ~~states~~ otherwise, the table compressor 100 always uses MaxIndependentSet for CaRT-selection and the integrated pruning and building algorithm for constructing regression trees.

**Experimental Results****Effect of Error Threshold on Compression Ratio**

(2) Please replace paragraph 133 on page 57 with the following paragraph showing changes.

Another crucial difference between fascicle and CaRT-based compression is that, when fascicles are used for compression, each tuple and as a consequence, every attribute value of a tuple is assigned to a single fascicle. However, in the table compressor 100, a predictor attribute and thus a predictor attribute value (belonging to a specific tuple) can be used in a number of different [[,]] CaRTs to infer values for multiple different predicted attributes. Thus, CaRTs offer a more powerful and flexible model for capturing attribute correlations than fascicles.

(3) Please replace paragraph 135 on page 58 with the following paragraph showing changes.

The compression ratios for the table compressor 100 are even more impressive for larger values of error tolerance (e.g., 10%) since the storage overhead of CARTs + outliers is even smaller at these higher error values. For example, at 10% error, in the compressed Corel data set, CaRTs consumer only 0.6 MB or 5.73% of the original table size. Similarly, for Forest-cover, the CaRT storage overhead reduces to 2.84 MB or 3.72% of the uncompressed table. The only exception is the Census data set where the ~~decrease~~ decreased in storage overhead is much steeper for fascicles than for CaRTs. One possible reason for the preceding is because of the small attribute domains in the Census data that cause each fascicle to cover a large number of tuples at higher error threshold values.

Effect of Random Sample Size on Compression Ratio